

# Queen's in the Global Knowledge Commons

Data Day 2016, Queen's University

Martha Whitehead, Vice-Provost and University Librarian

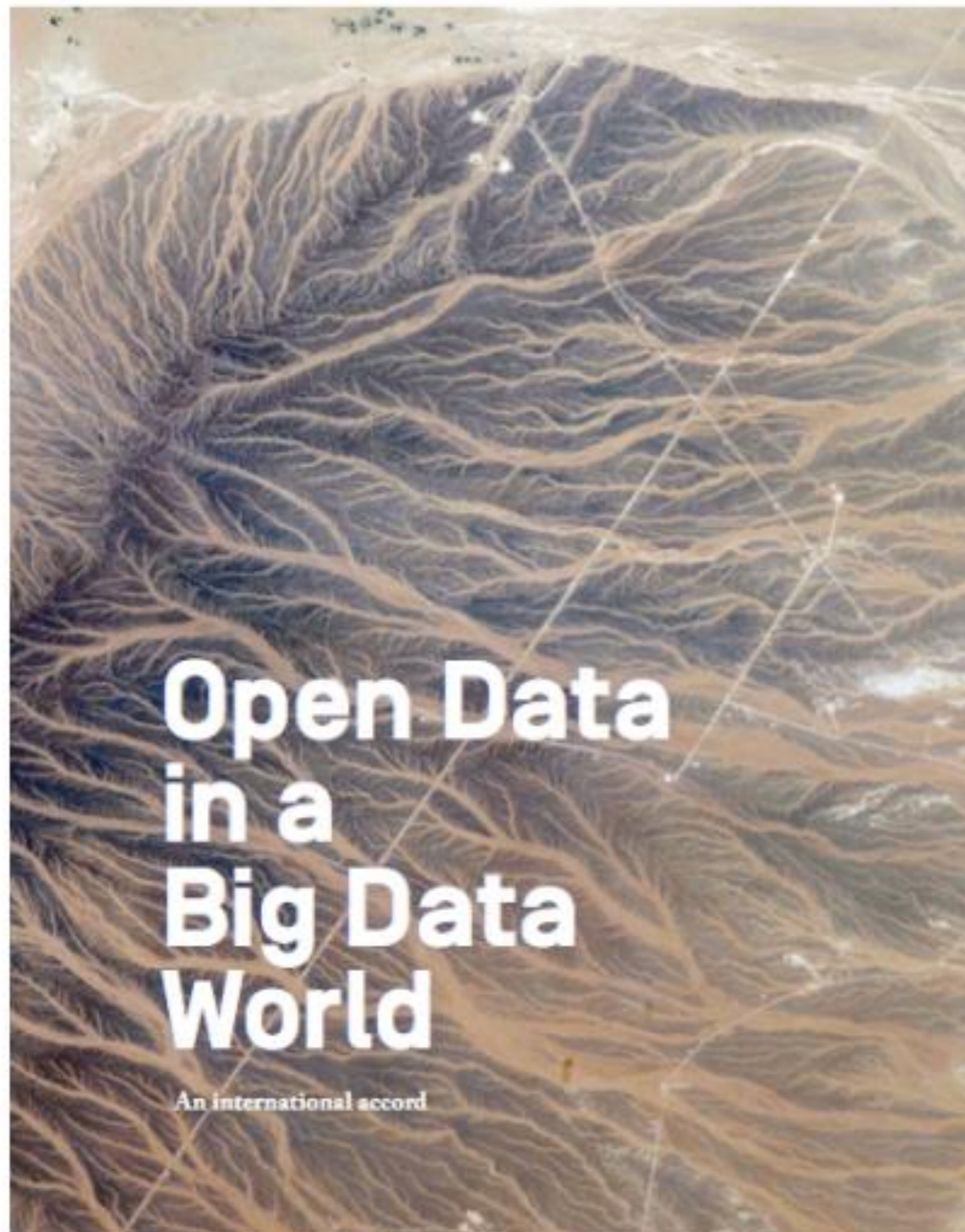
# Outline

## **What this is really about: research quality and impact**

- Open Science context of research data management
- Broadening the concept of scholarly communications
- Institutional RDM services emerging at Queen's
- National context and Portage Network opportunities
- International perspectives

# My Lens: Information Access

- ✓ Effective and sustainable scholarly communication
- ✓ Public policy that enables information access
- ✓ Respect for intellectual property
- ✓ Protection of privacy
- ✓ Preservation

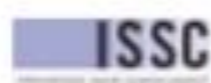
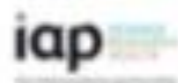


# Open Data in a Big Data World

An international accord

“Libraries, archives and repositories have a responsibility for the development and provision of services and technical standards for data to ensure that data are available to those who wish to use them and that data are accessible over the long term.”

Jointly endorsed by ARL,  
CARL, CAUL, RLUK and  
LIBER



# Merton's Scientific Norms (1942)

**Communalism:** scientific results are the common property of the community.

**Universalism:** all scientists can contribute to science regardless of race, nationality, culture, or gender.

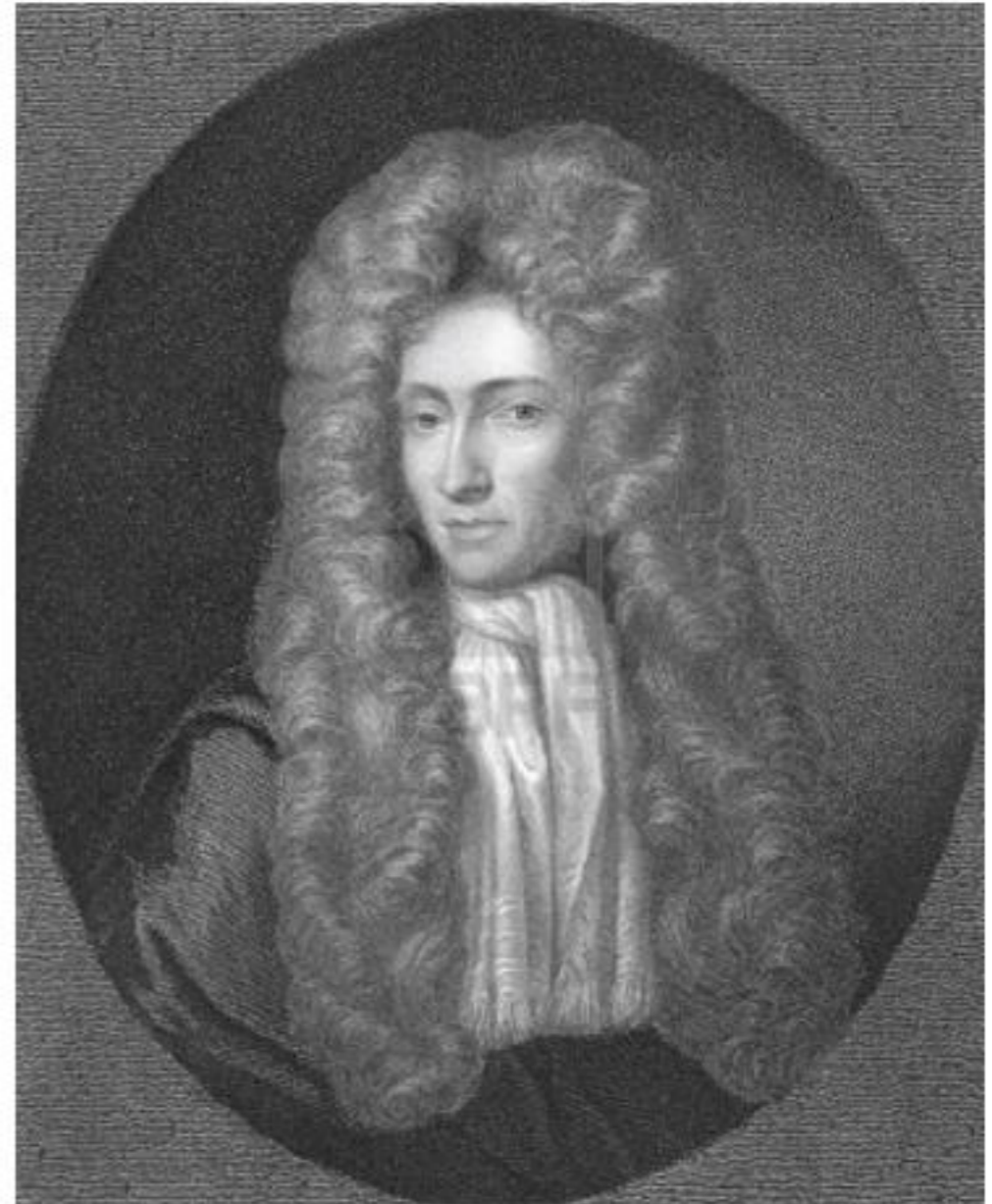
**Disinterestedness:** act for the benefit of a common scientific enterprise, rather than for personal gain.

**Originality:** scientific claims contribute something new

**Skepticism:** scientific claims must be exposed to critical scrutiny before being accepted.

# Skepticism -> Reproducibility

- Skepticism requires that the claim can be independently verified,
- This in turn requires transparency in the communication of the research process.
- Instantiated by Robert Boyle and the Transactions of the Royal Society in the 1660's.



# Really Reproducible Research

“Really Reproducible Research” (1992) inspired by Stanford Professor Jon Claerbout:

“The idea is: An article about computational science in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship. The actual scholarship is the complete ... set of instructions [and data] which generated the figures.” David Donoho, 1998





# Research Data Management Services

Library, guided by working group including University Research Services representatives, works closely **with researchers** to help with

- *Planning for data management*  
data management plans and grant applications
- *Preparing your data*  
coding, formatting, anonymizing, and more
- *Documenting your data*  
describing your data so others can understand and use it
- *Archiving your data*  
finding an appropriate repository or archive
- *Making your data discoverable*  
promoting open access, indexing, assigning DOI's

## Quick Links

- [Our Elbow Lake Environmental Education Centre](#)
- [Opinicon Natural History Blog](#)
- [QUBS Dataverse Data Archive](#)
- [Dispatches from the Field Blog](#)
- [Fowler Herbarium](#)
- [Field Biology at Queen's](#)
- [Queen's University](#)
- [Department of Biology](#)

## Field safety tips



Check Us Out on Facebook



Follow: QUBS Director Stephen Loughheed



Explore our Flickr Photostream



See our YouTube Channel

## News

MAY [Research and teaching ramping](#)

## Queen's University Biological Station

The Queen's University Biological Station (QUBS) is one of the premier scientific field stations in Canada. For almost 70 years, researchers and students have gathered at QUBS to conduct leading-edge research and participate in courses spanning ecology, evolution, conservation, geography, and environmental science.



*Getting dirty in Greg Bulte's ecology field course*

# Scholars Portal Dataverse

A Service of the Ontario Council of University Libraries

## Scholars Portal Dataverse

A Service of the Ontario Council of University Libraries



POWERED BY THE **Dataverse Network** PROJECT v. 3.6

## SP Dataverse Network



[Create Account](#)

[Log In](#)

[Advanced Search](#) [Tips](#)

The Scholars Portal Dataverse network is a repository for research data collected by individuals and organizations associated with Ontario universities. The Dataverse platform makes it easy for researchers to deposit data, create appropriate metadata, and version documents as you work. Access to data and supporting documentation can be controlled down to the file level, and researchers can choose to make content available publicly, only to select individuals, or to keep it completely locked. All data is hosted on Canadian servers, in a secure environment that conforms to industry best practices for maintaining data integrity and longevity.

For more information on Scholars Portal Dataverse, visit [Dataverse at Scholars Portal](#).

## Dataverses

**61** Dataverses

**i** A **Dataverse** is a container for research data studies, customized and managed by its owner.

### RECENTLY RELEASED DATAVERSES

<a href="#">Carleton University, MacOdrum Library Data Repository</a>	Aug 12, 2015
<a href="#">Wilfrid Laurier University Library Research Data Repository</a>	Aug 7, 2015
<a href="#">Public Data Collection</a>	Jul 20, 2015
<a href="#">Open Education Dataverse</a>	Jul 5, 2015
<a href="#">Steven Liss</a>	May 21, 2015

[View More >](#)

## Studies

**368** Studies, **4,861** Files, **16,469** Downloads

**i** A **study** is a container for a research data set. It includes cataloging information, data files and complementary files.

### RECENTLY RELEASED STUDIES

<a href="#">The Charter of Rights and Freedoms Survey, 2002 [Canada] by Centre for Research and Information on Canada</a>	Aug 13, 2015
<a href="#">Canadian Career College Students Survey, 2008 [Canada]: Phase I, II and III by R.A. Malatest &amp; Associates Ltd.</a>	Aug 13, 2015
<a href="#">Post-Secondary Education Drivers, 2004 [Canada] by Canada Millennium Scholarship Foundation</a>	Aug 13, 2015
<a href="#">Survey of Secondary School Students, 2004 by The Canada Millennium Scholarship Foundation</a>	Aug 11, 2015
<a href="#">Student Financial Survey, 2001-2002 [Canada] by Canada Millennium Scholarship Foundation</a>	Aug 11, 2015

# Ontario Library Research Cloud



- Established to provide cost-effective long-term storage for digital assets
- Large-scale cloud storage service using low-cost technologies
- Storage nodes at multiple Ontario campuses
- Advanced text mining tools
- Utilizes industry-standard APIs and interfaces to maximize compatibility with existing library use cases, including institutional repositories





## Accelerating the pace of Canadian Research

The **Centre for Advanced Computing** (formerly HPCVL) at Queen's University is a consortium comprised of Carleton University, University of Ottawa, and the Royal Military College of Canada. We specialize in secure, advanced computing resources and support for academic and medical researchers. CAC operates a high performance data centre as part of the Compute Canada family.



# Institutional concerns

- How can we ramp up to support funder guidelines for data management?
- Where is the infrastructure? (technical and human)
- Isn't this a national / international concern?
- Will researchers engage with RDM issues?

# Bigger than any one institution

- Resources vary by institution and region, and nobody has it solved completely
- Managing research data is complex (diversity of format, metadata, analytical tools): researchers and service providers need help
- Researchers and their data cross institutional, regional and disciplinary boundaries
- Preservation storage is not yet addressed

*Research data management is an element  
of national digital infrastructure*

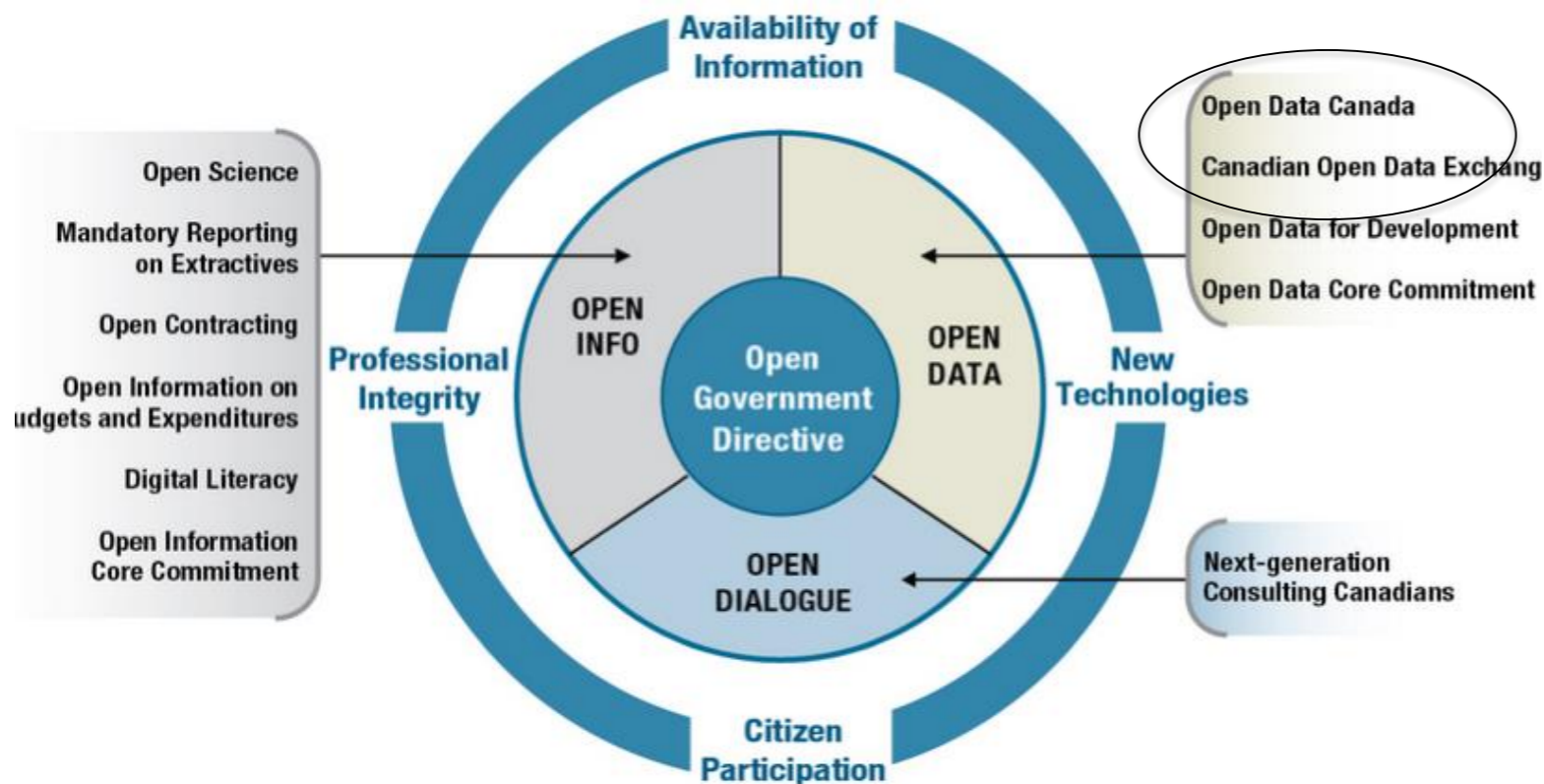
# National Context

- Policy
- Portage Network:
  - In Canada's RDM ecosystem
  - Its RDM framework:
    - Network of expertise
    - Infrastructure platforms and tools



# Open Science

## Canada's Action Plan on Open Government 2014-2016



*“...the Government of Canada will establish a government-wide approach to Open Science to increase access to federally-funded scientific publications and data.”*

### Tri-Agencies (CIHR, NSERC, SSHRC)

- Consultation in 2013 (Report: *Capitalizing on Big Data: Toward a Policy Framework for Advancing Digital Scholarship in Canada*)
- *Tri-Agency Statement of Principles on Digital Data Management*



**CARL  
ABRC**

**portage**

SERVICES PARTAGÉS POUR LES DONNÉES DE RECHERCHE  
SHARED STEWARDSHIP OF RESEARCH DATA

Portage slides courtesy of Chuck Humphrey

# What is Portage?

A national, library-based research data management network that coalesces initiatives in research data management to build capacity and to coordinate activities better.

- 2013 (Dec): Concept meeting at the ARC Hotel
- 2014 (Mar): Launch of Project ARC, a one-year pilot
- 2015 (May): Framework and transition plan for operating network
- 2015 Sep – 2017 Aug: Portage Director developing services, governance and operational models
- 2017 Aug:

# Portage in Canada's RDM ecosystem

# Portage Principles

- Research data are a public good
- Intelligent access: openness, with respect for privacy
- Collaborative approaches: cost savings and shared expertise
- Inclusiveness: aim to serve all researchers and create a more level playing field
- Commitment to standards and interoperability
- International relationships: liaise internationally and ensure Portage keeps pace with international practices
- Respect for differences: flexibility to meet the needs of different regions, institutions, and disciplines
- Open source: Tools will be contributed back to the community
- Stewardship: a sense of responsibility for managing research data over the long term

# Other RDM principles

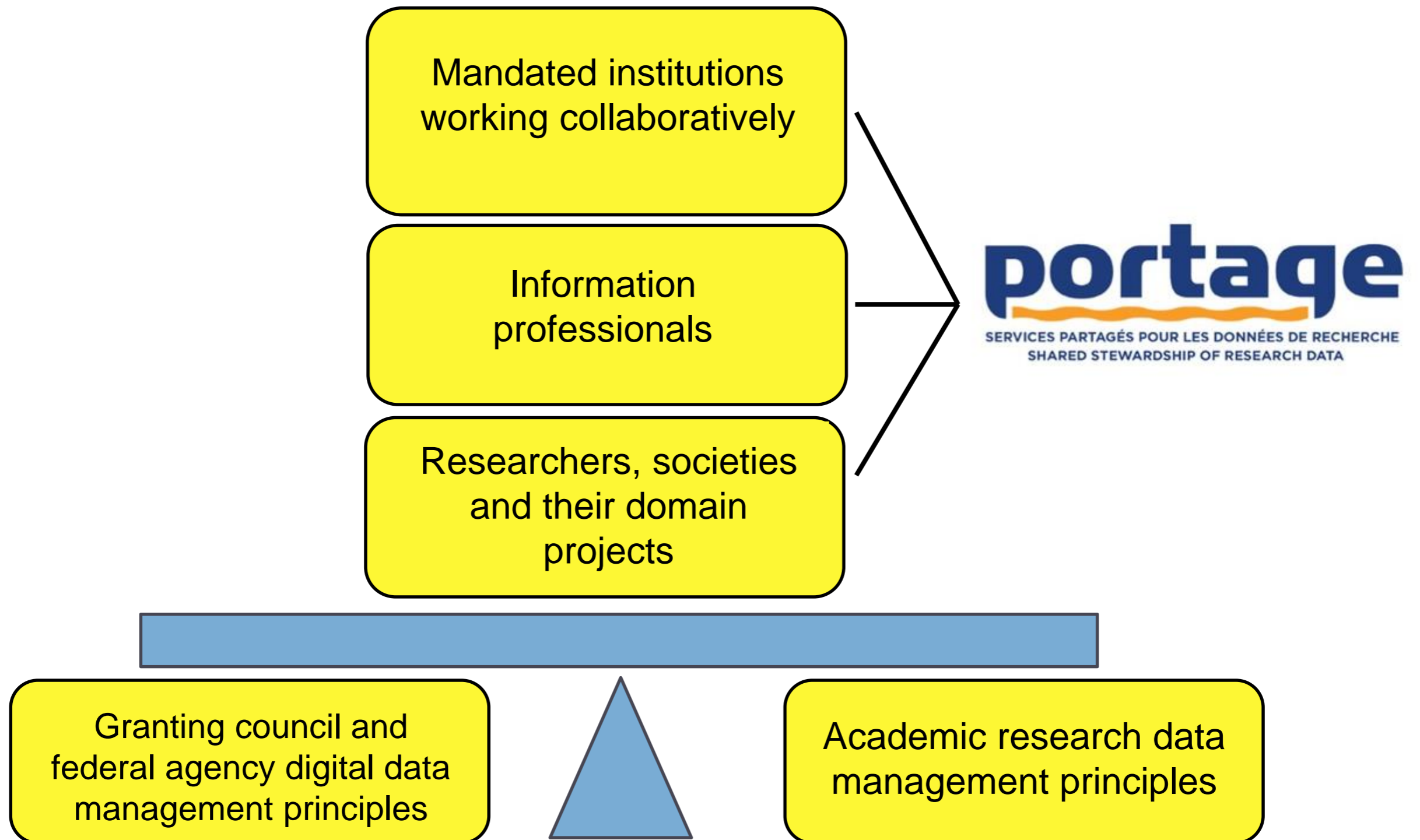
## Tri-agency Statement of Principles on Digital Data Management

- Canadian Institutes of Health Research
- Natural Sciences and Engineering Research Council of Canada
- Social Sciences and Humanities Research Council of Canada
- January 2016

## Research Data Management in Canadian Universities: A statement of principles

- Research Data Canada and Vice-Presidents of Research from the U15
- March 2016

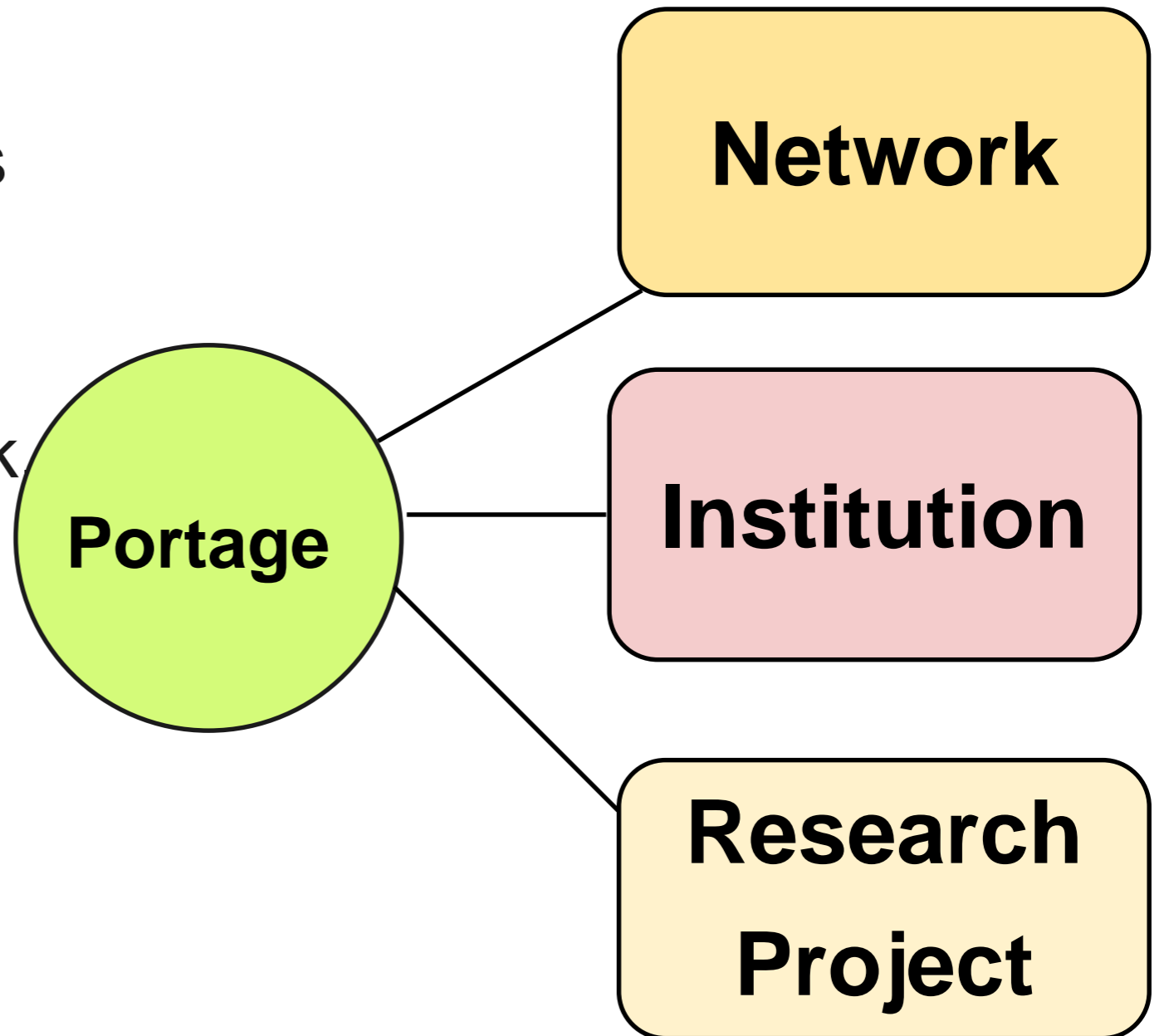
# Bottom-up approach



# Is there a best level to collaborate?

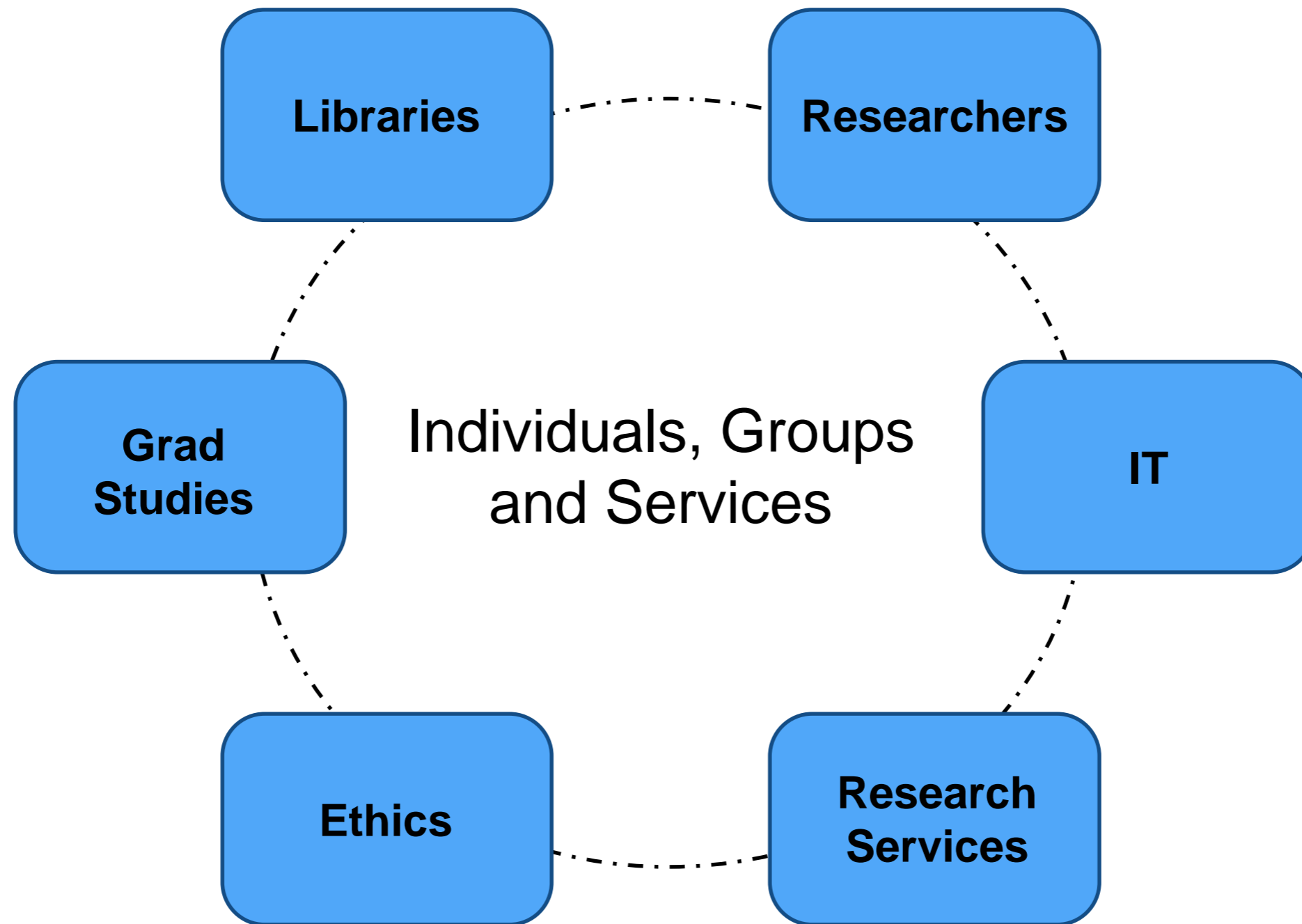
Portage aims to provide infrastructure and services at the levels of the research project, the institution, and the network.

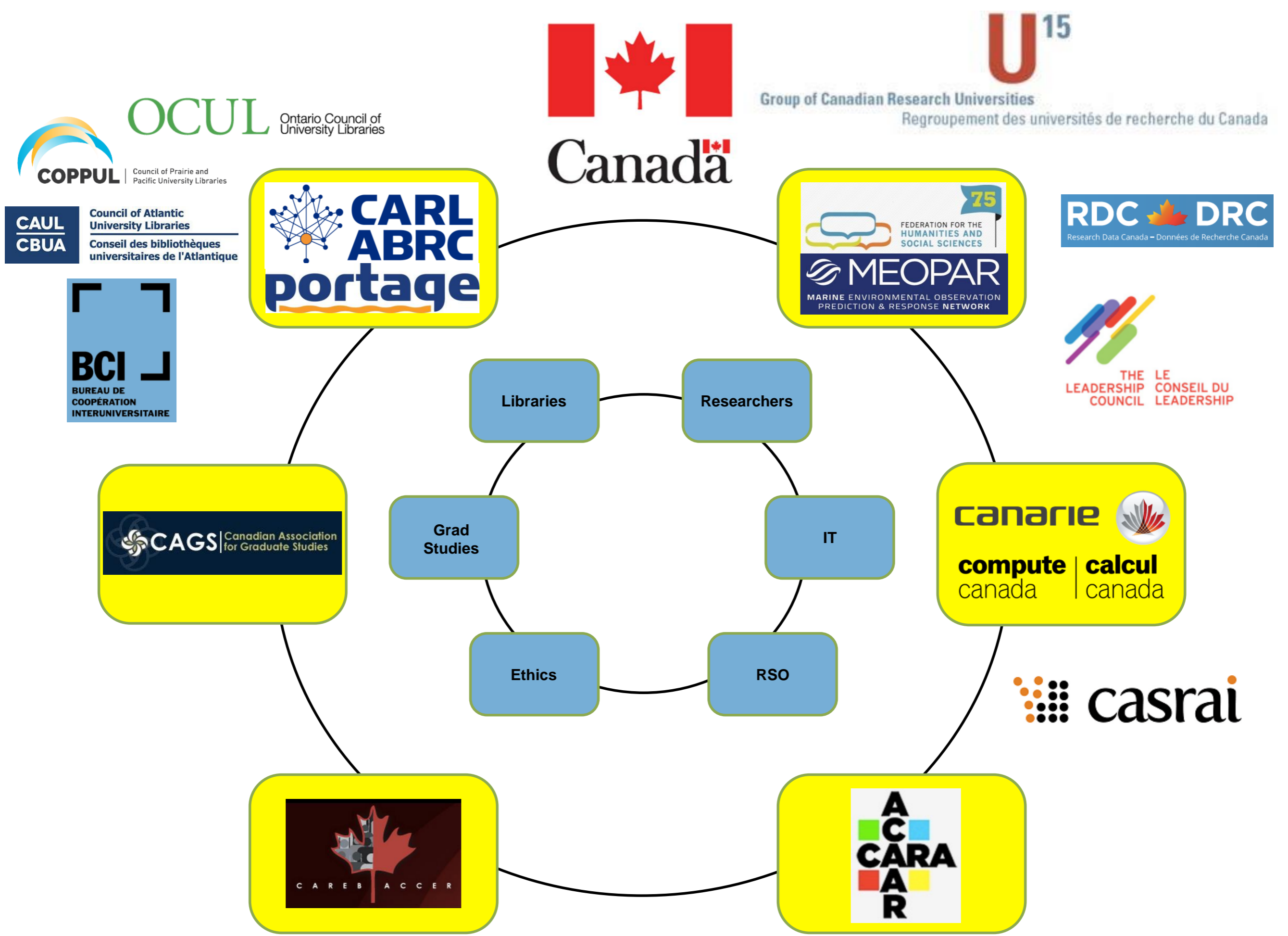
By network, we mean collaborating in regional and national partnerships.





# Shared stewardship at the institutional level





**OCUL** Ontario Council of University Libraries

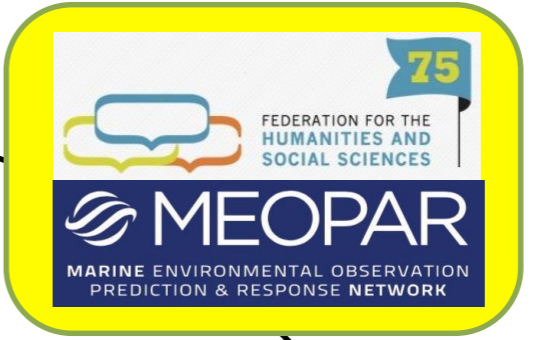


**Group of Canadian Research Universities**  
Regroupement des universités de recherche du Canada

**U15**

**COPPUL** Council of Prairie and Pacific University Libraries

**CAUL CBUA** Council of Atlantic University Libraries  
Conseil des bibliothèques universitaires de l'Atlantique



**RDC DRC** Research Data Canada - Données de Recherche Canada

**BCI** BUREAU DE COOPÉRATION INTERUNIVERSITAIRE

**THE LEADERSHIP COUNCIL / LE CONSEIL DU LEADERSHIP**

**Libraries**

**Researchers**



**Grad Studies**

**IT**



**Ethics**

**RSO**

**casrai**



# Portage RDM Framework

# Portage Network Research Data Management Framework

## RDM Framework

Data Planning and Production	Data Deposit	Data Curation	Data Preservation	Data Access	Reuse & Integration
DMPs & Project-level research data management	Staging & institutional repositories	Trustworthy data repositories		Institutional and domain repositories	Virtual research environments

## Portage Network of Experts

Data Management Plans	Data Curation	Data Preservation	Data Discovery
Training			

## RDM Tools & Platforms

Data Management Plans	File Transfer and Deposit	Curation Toolkit and Preservation Pipeline	Data Discovery and Data Integration Toolkit
-----------------------	---------------------------	--	---

## Portage Tools

DMP Assistant	DVN2Arch Islandora2Arch Globus FTP	Open Science Framework FPR-PREFER, Archivematica Globus Publication	Federated Search (UBC)
---------------	--	---	------------------------



# Portage Network of Expertise

Capitalize on the research data management expertise and services in libraries.

- Expert Groups currently operational
  - Data Management Plan Expert Group
  - Data Preservation Expert Group
  - Data Discovery Expert Group
  - RDM Learning Expert Group
- Expert Groups forthcoming
  - Data Curation Expert Group
  - Ethics and Data Expert Group



Shared stewardship of research data

[Version française](#)

Signed in as [mw9@queensu.ca](#)

[My plans](#)

[Create plan](#)

[About](#)

[Help](#)

Plan was successfully created.

## My plan (Portage Template)

**Plan details**

**Portage Data Management Questions**

**Share**

**Export**

Please fill in the basic project details below and click 'Save' to save

[Cancel](#)

Plan name

My plan (Portage Template)

ID

Grant number

Principal Investigator/Researcher

Principal Investigator/Researcher ID

Plan data contact

Description

# Infrastructure platforms and tools

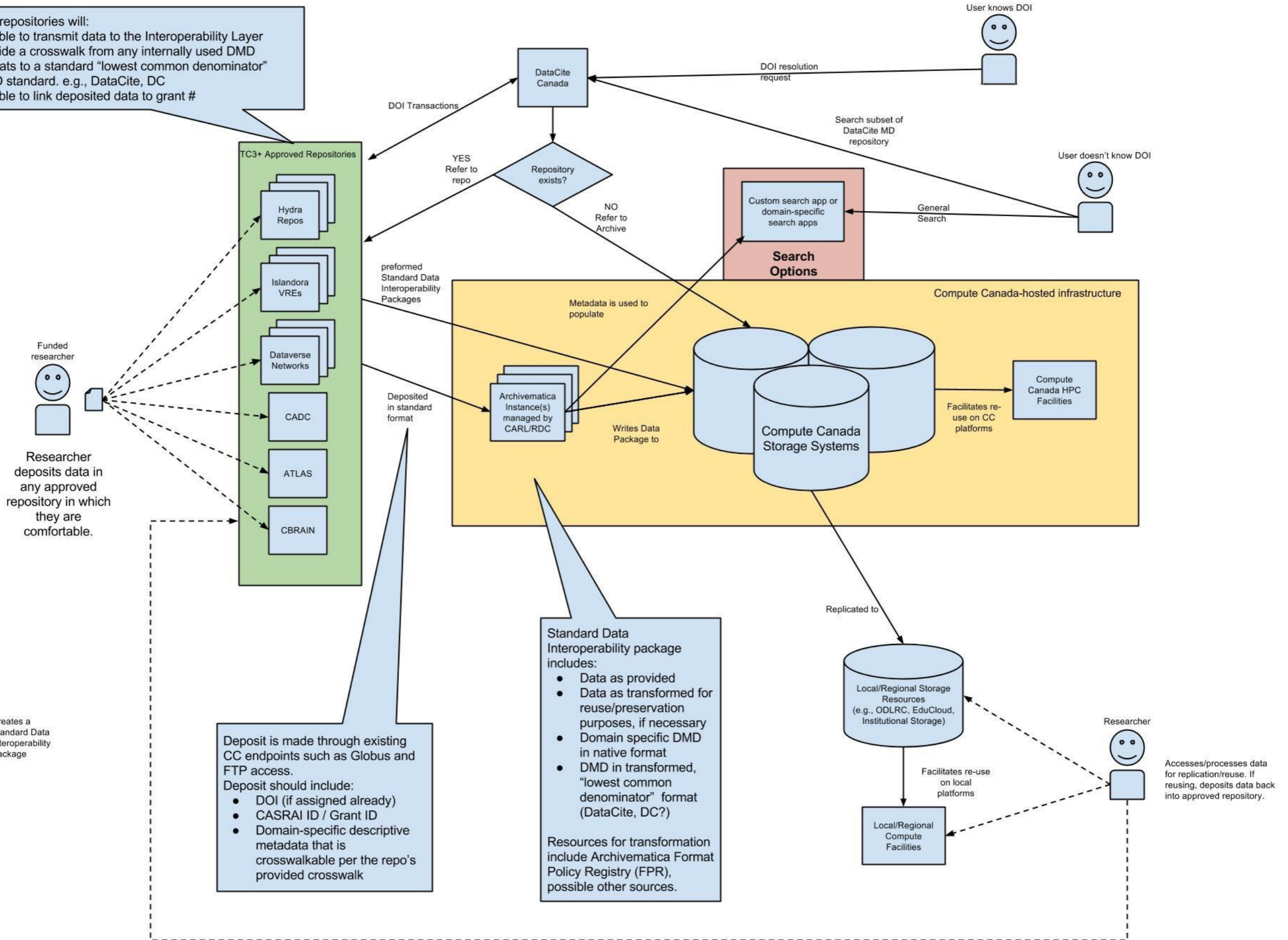
Develop RDM platforms and tools through partnerships and memoranda of understanding.

- Current partnerships in place or nearly signed
  - DMP Assistant (UAL-CARL)
  - National Research Data Repository Prototype (CC-CARL)
  - Preservation Pipeline (CC-OCUL/SP-UTL-CARL)
  - Dataverse-to-Archivematica middleware (OCUL/SP-CARL)
- Forthcoming
  - PREFER: Policy Format Register
  - Federated data discovery system
  - Data curators toolkit

# Preservation Workflow Concept

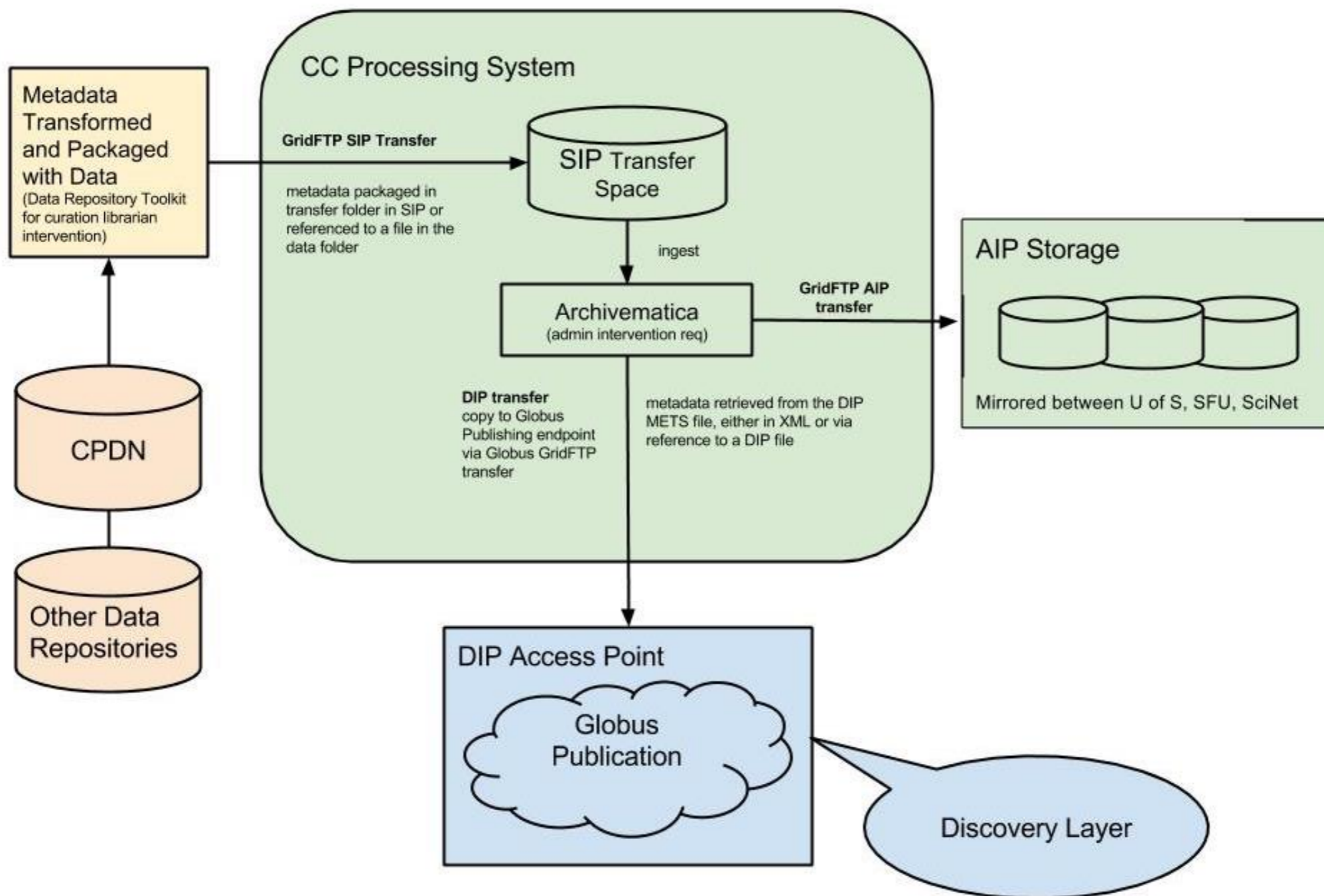
Approved repositories will:

- be able to transmit data to the Interoperability Layer
- provide a crosswalk from any internally used DMD formats to a standard "lowest common denominator" DMD standard. e.g., DataCite, DC
- be able to link deposited data to grant #





# National Preservation Pipeline Service with Compute Canada (in progress)



- CARL/Portage and CC: jointly manage the development through shared planning processes
- CARL/Portage: research data management expertise and in-kind contributions
- CC: project management and software development expertise and necessary computational resources
- all derived software products will be released to the wider research community through an *open source licence* (except software components from other parties that are covered by their own licensing regime)
- non-exclusive relationship

# Stakeholder engagement

Build a community of practice for research data management through well established working relationships among stakeholders across research sectors.

- Regional library consortia
- Individual academic libraries
- Academic RDM stakeholders, e.g., VPR's, REBs, Scholarly Associations, Networks of Centres of Excellence
- Other RDM stakeholders, e.g., Research Data Canada, Leadership Council for Digital Infrastructure



**THE PORTAGE NETWORK** is dedicated to the shared stewardship of research data in Canada through:

- Developing a national research data culture
- Fostering a community of practice for research data
- Building national research data services and infrastructure

Launched in 2014 by the [Canadian Association of Research Libraries](#), Portage works within the library community to coordinate expertise, services, and technology in research data management, seeking to collaborate with other research data management stakeholders.

**Research data culture** represents widely shared values and principles for digital data management.

**A community of practice for research data** consists of stakeholders working collaboratively to ensure data are accessible to address complex research issues.

Research data management activities span sectors, domains, and jurisdictions,



## DMP Assistant

[Sign In](#)[Sign Up](#)

If you have an existing account with DMP Assistant or previous version of DMP Builder.

New to DMP Assistant? Sign up today.

## NEWS

New website for the Portage Network

APRIL 18, 2016

Compute Canada and CARL Join Forces to Build a National Research Data Platform

JANUARY 27, 2016

[Download](#)

Tweets by [@portageCARLABRC](#)

[Portage](#) [@portageCARLABRC](#)

You can now discover [@portageCARLABRC](#) on the web! [portagenetwork.ca](#) is the new home for shared stewardship of research data in Canada

[Embed](#)[View on Twitter](#)

# International Context

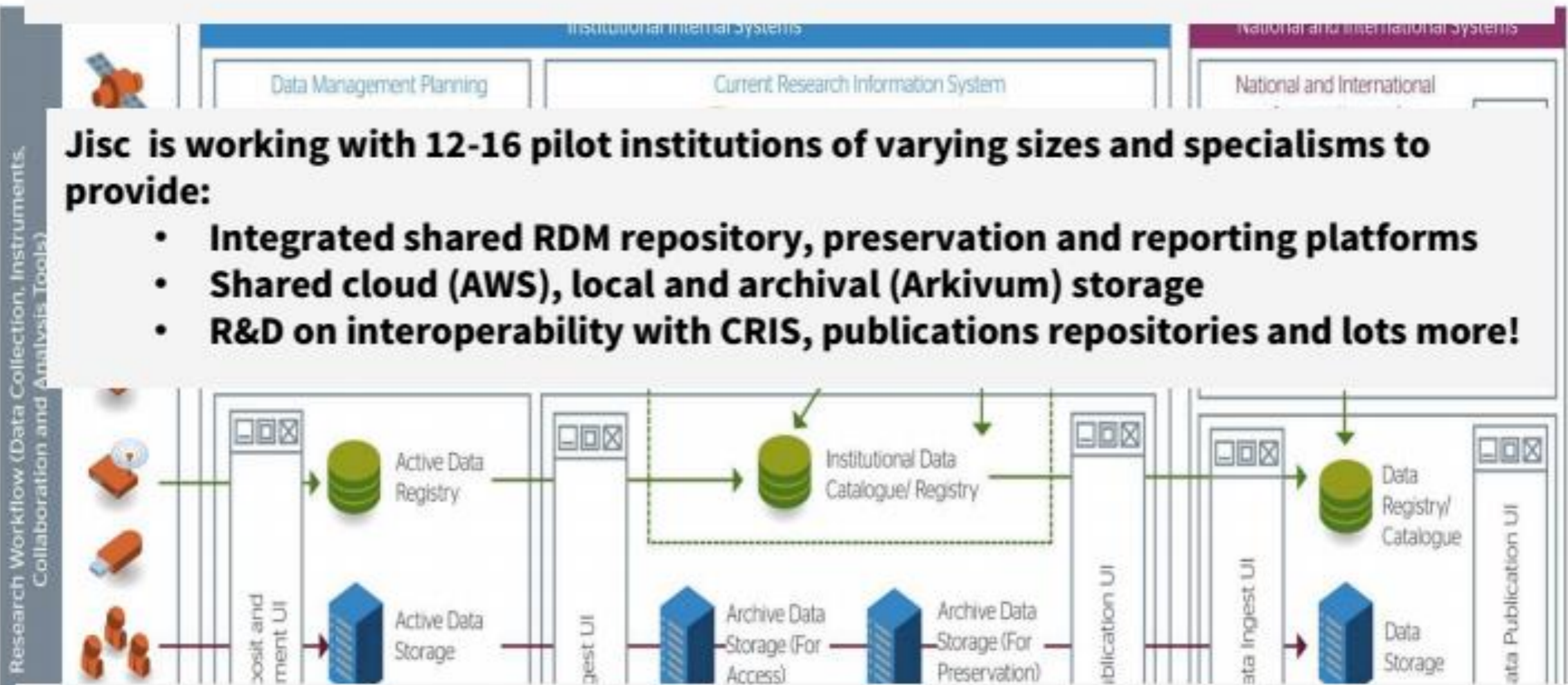
- ✧ National services, but also key international issues:
  - ✧ New modes of scholarly communications
  - ✧ Open science framework
  - ✧ Interoperability for web-based scholarship
  - ✧ Open repositories

With thanks to Kathleen Shearer, Executive Director of COAR, and participants at IDCC16, RLUK 2016, CNI Spring 2016 Meeting, COAR General Assembly 2016, CARL/ARL Spring 2016 Joint Meeting

**UK Universities must comply with funder mandates to manage RDM through the lifecycle e.g. EPSRC**

**Jisc is working with 12-16 pilot institutions of varying sizes and specialisms to provide:**

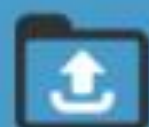
- **Integrated shared RDM repository, preservation and reporting platforms**
- **Shared cloud (AWS), local and archival (Arkivum) storage**
- **R&D on interoperability with CRIS, publications repositories and lots more!**



**Currently in procurement. Framework of Suppliers to be decided by May 2016  
Business case for production service by December 2017**



DANS promotes **sustained access** to digital research data files and encourages researchers to **archive** and **reuse** data.



### ARCHIVING

Deposit your datasets in EASY or send research data and publications to NARCIS.

DEPOSIT



### REUSE

Find datasets, publications, researchers, projects and institutions via NARCIS and EASY.

SEARCH



### TRAINING & CONSULTANCY

Let DANS advise you on data management and certification of digital archives.

ADVICE FROM DANS

#### SPOTLIGHT

## Collaboration Mendeley and DANS on archiving datasets

DANS

MENDELEY DATA

DANS and Mendeley are pleased to announce their collaboration in the long-term archiving of datasets that are published via the new Data Repository platform, Mendeley Data.

Together we provide open, persistent, certified and stable access to digital research data.

Mendeley launched a new data repository platform: Mendeley Data ([data.mendeley.com](http://data.mendeley.com)). The data submitted by researchers on this platform will be included in the long-term archive of DANS. This collaboration will allow DANS to store even more digital research data in the long term.

#### NEWS

### News items in English

Please note that not all news items are translated into English. Please visit the Dutch page of this website for all news items. If you have any questions, please contact us via [info@dans.knaw.nl](mailto:info@dans.knaw.nl).

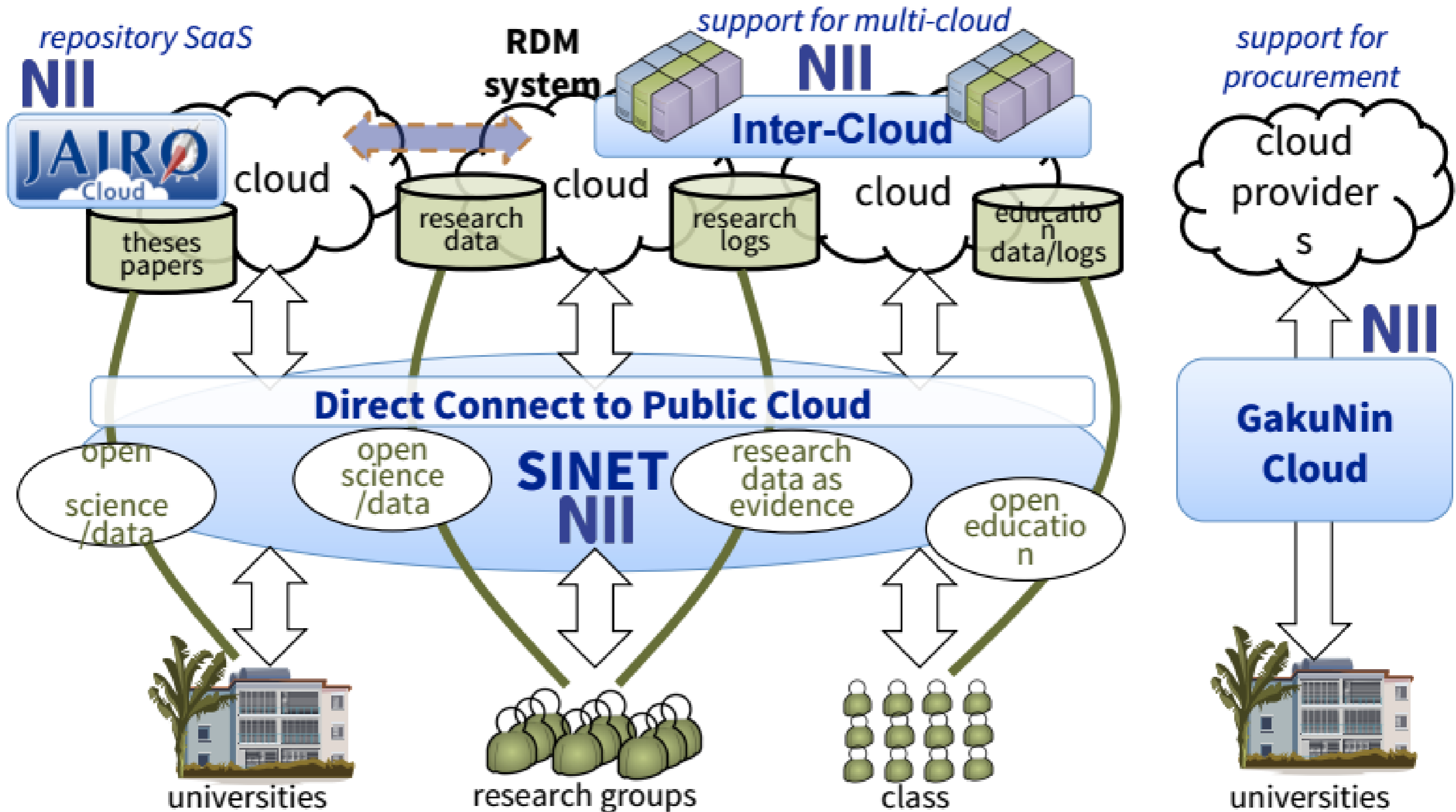
»»

### DANS facilitates open access data

From now on, depositors can make their datasets available under full open access in EASY. Interested parties can download this type of dataset without registration. This makes research data sharing even easier.

»»

# National Institute of Informatics, Japan Academic Cloud Services



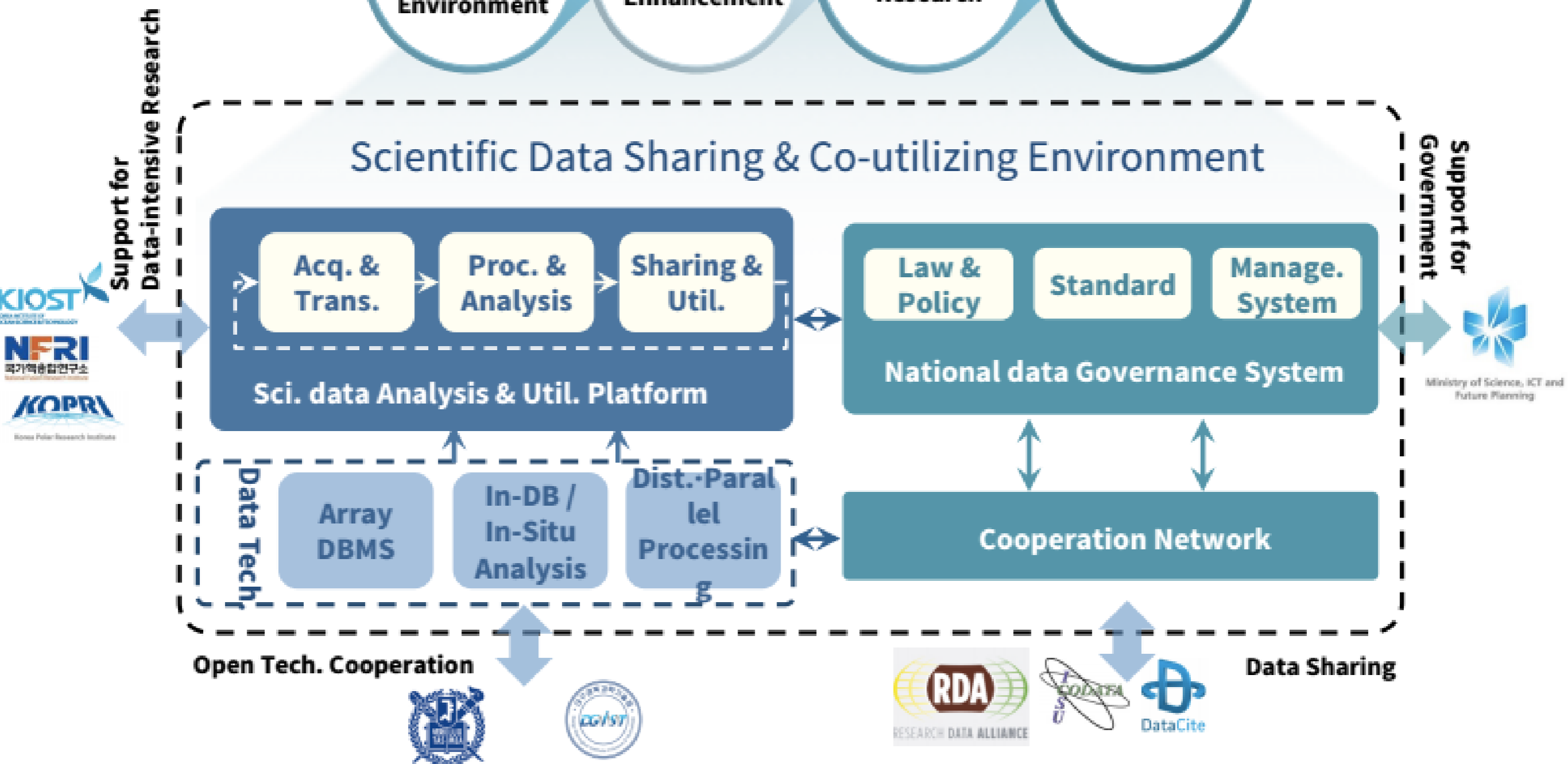


# Scientific Data Sharing & Co-utilization

## KISTI, South Korea



Provision of Solutions for Social Issues through Digital Science



# National Data Service (US)

The National Data Service (NDS) is an emerging vision for how scientists and researchers across all disciplines can find, reuse, and publish data. It builds on the data archiving and sharing efforts already underway within specific communities and links them together with a common set of tools designed around the following capabilities:

## Search

The NDS will allow users to easily search for data across disciplinary boundaries. As users hone in on data of interest, they can easily switch to discipline-specific tools.

## Publish

The NDS will connect users to tools for building and sharing collections of data. It will help users find and deliver data to the best repository for data-publishing.

## Link

The NDS will create robust connections between data and published articles. When researchers reference an article, they have ready access to the underlying data.

## Reuse

The NDS will not only provide access to data for download, it will provide tools for transferring data to processing platforms or allow analysis to be attached to the data.



# Barend Mons



Biosemantics Group LUMC and EMC



LS integrator Netherlands eScience Center



Chair of DTL-data  
Head of ELIXIR node NL



EC member of Open PHACTS

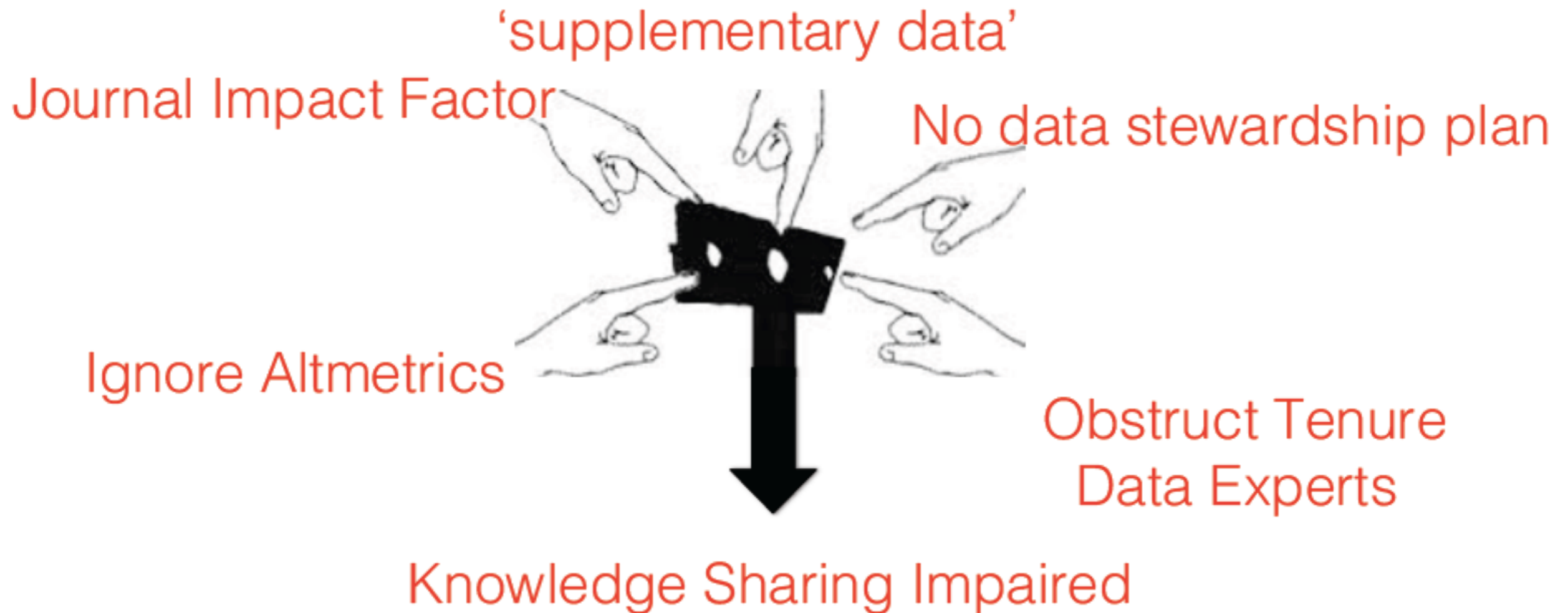
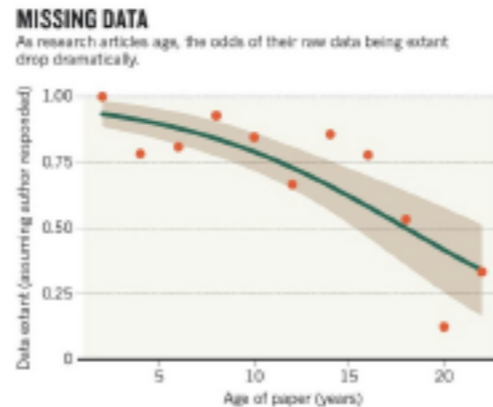


Chair of High Level Expert Group EOSC



The opinions expressed in this presentation are my personal opinions and do not necessarily reflect the draft report of the High Level Expert Group for the European Open Science Cloud.

# Malpractices.....



## EOSC: ‘Internet approach’

- **Minimal** international guidance and governance
- **Maximum** freedom to implement.
- **Globally** interoperable and accessible
- **Globally** embedded in a ‘**Commons**’



## EOSC: **Key requirements**

- **New modes of scholarly communication**
- **Modern reward and recognition practices need to support data sharing and re-use**
- **Innovative, fit for purpose funding schemes for sustainable underpinning infrastructures**
- **Core data experts need to be trained and their career perspective significantly improved**
- **Cross-disciplinary collaboration-specific measures for review, funding and infrastructure**
- **Support for the transition from scientific insights towards societal innovation**
- **The EOSC needs to be developed as an eco-system of infrastructures**
- **Key Performance Indicators should be developed for the EOSC**
- **The EOSC should enable automation of data processing and thus machine actionability is key.**
- **FAIR principles**

**GROUP MENU**[Group Home](#)[Members](#)[Links & Files](#)[Google Forum](#)[Calendar](#)[Subscribe to group](#)**GROUP LEADER**

Barend Mons



31 Member(s)

[FORCE11](#) » [Groups](#) » [FAIR Data Publishing Group](#)**FAIR DATA PUBLISHING GROUP**[To view and comment on FAIR principles click here](#)

In the eScience ecosystem, the challenge of enabling optimal use of research data and methods is a complex one with multiple stakeholders: Researchers wanting to share their data and interpretations; Professional data publishers offering their services, software and tool-builders providing data analysis and processing services; Funding agencies (private and public) increasingly concerned with proper Data Stewardship; and a Data Science community mining, integrating and analysing the output to advance discovery. Computational analysis to discover meaningful patterns in massive, interlinked datasets is rapidly becoming a routine research activity. Providing machine-readable data as the main substrate for Knowledge Discovery and for these eScientific processes to run smoothly and sustainably is one of the Grand Challenges of eScience.

This groups main aim is to create and put up for community endorsement a document that is a general 'guide to FAIRness of data', not a "specification".

- Data should be Findable
- Data should be Accessible
- Data should be Interoperable
- Data should be Re-usable.



# Strategies for improving openness and reproducibility

Andrew Sallans

Center for Open Science

<http://cos.io/>

@OSFramework



ALFRED P. SLOAN  
FOUNDATION

JOHN TEMPLETON FOUNDATION

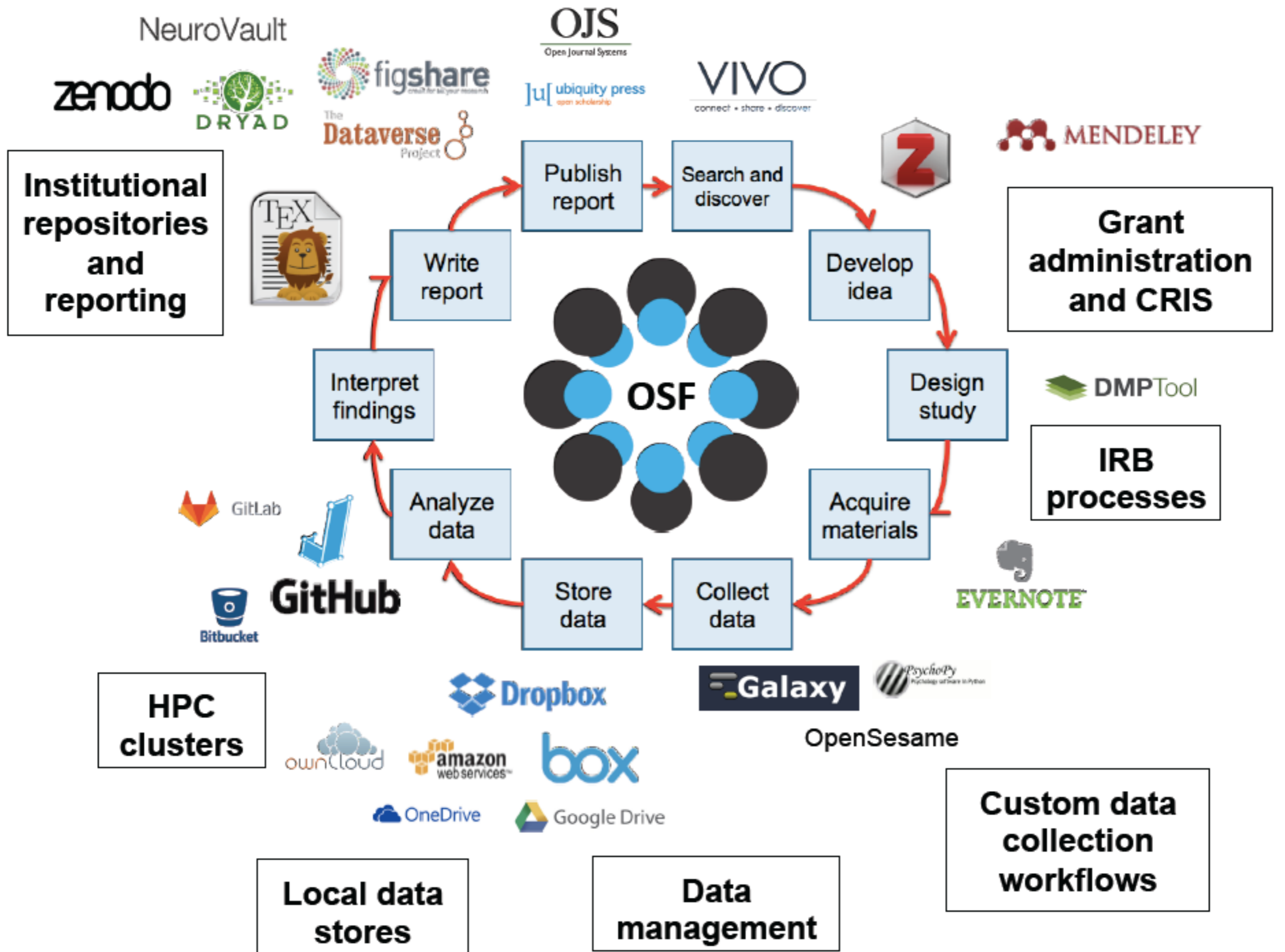
SUPPORTING SCIENCE ~ INVESTING IN THE BIG QUESTIONS



Laura and John Arnold Foundation®







# Establishing New Levels of Interoperability for Web-Based Scholarship

Herbert Van de Sompel  
Los Alamos National Laboratory  
[@hvdsomp](https://twitter.com/hvdsomp)



Cartoon by:  
Patrick Hochstenbach

Acknowledgments: Michael L. Nelson,  
David Rosenthal, Geoff Bilder, Simeon  
Warner, Harihar Shankar, Shawn Jones



Herbert Van de Sompel  
COAR Annual Meeting, Vienna, Austria, 12/04/2016



## D-Lib Magazine

November/December 2015

Volume 21, Number 11/12

[Table of Contents](#)

---

## Reminiscing About 15 Years of Interoperability Efforts

Herbert Van de Sompel

Los Alamos National Laboratory

[herbertv@lanl.gov](mailto:herbertv@lanl.gov)

Michael L. Nelson

Old Dominion University

[mln@cs.odu.edu](mailto:mln@cs.odu.edu)

DOI: [10.1045/november2015-vandesompel](https://dx.doi.org/10.1045/november2015-vandesompel)

---

# Repository-Centric Interoperability Paradigm

Address interoperability challenges from the perspective of a node, e.g. an IR, a publisher, a web-based authoring portal, a software repository, ...

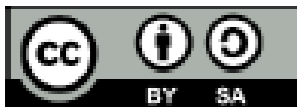
- **The node at the center of the universe**
- Define a machine interface for your node, expect others to use it
- Piggybacking on the web without truly embracing its core technologies
- The node resembles a brick & mortar library that can be visited subject to well-intended yet idiosyncratic policies – the interface



# Web-Centric, Resource-Centric Interoperability Paradigm

Address interoperability challenges from the perspective of the web

- **The resource at the center of the universe**
  - The notion of a node, a repository, not even of a web server exists in the architecture of the web
- The tools of the interoperability trade are the primitives of the web



# COAR Roadmap: Future Directions for Repository Interoperability

## Addressing Strategic Challenges

- Adoption of common metadata, identifiers (for authors, institutions, research funding organizations, publications), vocabularies and taxonomies
- Interaction with other local infrastructures and external systems (administrative research systems, publishing platforms, e-learning systems, etc)



## SESSION 2. Launch of the Open Peer Review Module for Institutional Repositories

**12:20-13:15 h.**

### **Presentation of the project and first outcomes**

Roundtable with members of the OPRM Consortium (Open Scholar, ARVO, DIGITAL.CSIC, Institute for Artificial Intelligence (CSIC-IIA), e-IEO, SECABA-University of Granada)

### **Open Peer Review Module for Open Access Repositories**

Pandelis Perakakis, Open Scholar.

Presentation available at [■ DIGITAL.CSIC](#)

### **The Open Peer Review Module... some technical details**

Emilio Lorenzo, ARVO Consultores.

Presentation available at [■ DIGITAL.CSIC](#)

### **Reputation in the Academic World**

Nardine Osman and Carles Sierra, Artificial Intelligence Research Institute (IIIA-CSIC).

Presentation available at [■ DIGITAL.CSIC](#)

### **OPRM Pilot Project in Digital.CSIC: first experience and thoughts**

Isabel Bernal, Digital.CSIC, institutional repository of the Spanish National Research Council (CSIC).

Presentation available at [■ DIGITAL.CSIC](#)

### **The Open Peer Review Module at the Spanish Institute of Oceanography (IEO)**

Concha Mosquera de Arancibia, e-IEO, institutional repository of the Spanish Institute of Oceanography.

Presentation available at [e-IEO](#)

# In conclusion...

## **Key Aims**

- ✧ Open global ecosystem of infrastructures
- ✧ New modes of scholarly communications
- ✧ Modern reward and recognition practices

*A global knowledge commons*



## **And key questions for us to ponder**

- ✧ What services are best offered locally, nationally, internationally?
- ✧ In developing services locally, how do we ensure international interoperability?
- ✧ What can we do to explore new modes of scholarly communications and modern reward and recognition practices?

Thank you!

Questions? Thoughts?